Assessment Resource Center
University of Missouri

College of Education
2800 Maguire Blvd.
Columbia, MO 65211-3290
PHONE (573) 882-4694
FAX (573) 882-8937

To: MO TAC

From: Paul Hirsch

Date: 11/24/2015

Subject: 2015 MAP-A (Science) Technical Report- Main Document

This note accompanies the draft 2015 MAP-A Science Technical Report.

The data and text in the report have been updated to refer to the 2014-2015 assessment year: see tables in the *Scoring and Reporting* and *Reliability and Validity* sections of the report. Mathematics and Communication arts were not assessed with the MAP-A portfolio for the 2014-2015 assessment year, no substantive changes have been made in the report's narrative.

Missouri Assessment Program-Alternate (MAP-A)

2015 Technical Report

DEVELOPED BY
Missouri Department of Elementary and Secondary Education
University of Missouri/Assessment Resource Center
Measured Progress

Contents

Contents (contd.)

Overview

The purpose of this report is to document the technical aspects of the 2014-2015 Missouri Assessment Program-Alternate (MAP-A) assessment. This was the ninth year of the MAP-A program in its current design. In the spring of 2015 students in grades 5, 8, and 11 participated in the MAP-A science assessment. Previous years students participated in the MAP-A as follows:

- Grades 3 & 4: Mathematics and communication arts;
- Grade 5: Mathematics, communication arts, and science;
- Grades 6 & 7: Mathematics and communication arts;
- Grade 8: Mathematics, communication arts, and science;
- Grade 10: Mathematics only;
- Grade 11: Communication arts and science.

Mathematics and communication arts MAP-A assessments were operational from 2006 through 2014. The science assessment for MAP-A was developed and piloted in 2007 and became operational in 2008. This report provides information about the technical quality of the mathematics, communication arts and science assessments, including a description of the processes used to develop, administer, and score the MAP-A, and how the scores are reported and analyzed.

Organization of the Report

The organization of this report is based on the conceptual flow of an assessment's life span. It begins with an overview of the initial test specifications and addresses all the intermediate steps that lead to final score reporting. The second section addresses the general design of the MAP-A, the ongoing development process, the specific designs of the communication arts, mathematics, and science assessments, the MAP-A format, and the administration of the assessment. The third section addresses scoring and reporting of MAP-A results. The fourth section addresses the reliability and validity of the MAP-A. The fifth section addresses security of MAP-A information. The report also includes references and appendices as appropriate.

This report describes several technical aspects of the 2015 MAP-A in an effort to contribute to the accumulation of validity evidence to support MAP-A score interpretations. Because it is the interpretations of scores that are evaluated for validity, not the assessment itself, this report presents documentation to substantiate intended interpretations (AERA, 1999). In the case of the MAP-A, however, construct validity is a major factor in score interpretation. The information in this report contributes important information to the validity assertion by addressing the following aspects of the MAP-A:

- Design and alignment with Missouri's standards;
- Administration;
- Scoring;
- Reporting;
- Achievement levels.

Purpose of the MAP-A

The Individuals with Disabilities Education Act (IDEA) requires that students with disabilities be included in each state's system of accountability and that students with disabilities have access to the general curriculum. The No Child Left Behind Act (NCLB) also speaks to the inclusion of all children in a state's accountability system by requiring states to report student achievement for all students as well as for groups of students on a disaggregated basis. These federal laws reflect an ongoing concern about equity. All students should be academically challenged and taught to high standards; all students should be involved in the educational accountability system.

To ensure the participation of all students in the state's accountability system, the Missouri Department of Elementary and Secondary Education (DESE) has developed the MAP-A. Only IDEA-eligible students with the most significant cognitive disabilities are expected to participate in the MAP-A. Students with moderate disabilities participate in the standard MAP Grade-Level and End-of-Course assessments, with appropriate accommodations.

The MAP-A is a portfolio-based assessment that measures student performance based on alternate achievement standards. The MAP-A is aligned with Missouri's Show-Me Standards, Grade Level Expectations (GLEs) and Alternate Grade Level Expectations (AGLEs) in communication arts, mathematics, and science. Missouri educators worked with DESE and its contractor, Measured Progress, to develop and review the AGLEs and to design the assessment blueprint for alternate assessment of eligible Missouri students.

MAP-A results are intended to inform stakeholders about student achievement on Missouri's communication arts, mathematics, and science standards and AGLEs. The results should be used for program and instructional improvement and as a component of school accountability.

The MAP-A assesses student performance on two Alternate Performance Indicators (APIs) in each of two content-area strands in communication arts and two content-area strands in mathematics. It also assesses performance on four APIs in science, which are selected from six strands. Teachers observe and assess a student's performance and collect evidence in each strand during two distinct collection periods. The assessment effectively links standards, curriculum, instruction, and assessment and is scored using three criteria: 1) level of accuracy, 2) level of independence, and 3) connection to the standards. The collected evidence provides documentation of a connection between the Show-Me Standards and instruction.

Development of the MAP-A

Considering the needs of Missouri's assessment programs at the time, among them efforts to ensure participation of all students in the state's accountability system, alignment of assessments with Missouri's Show-Me Standards and GLEs, and continued improvement to the state's assessment program, DESE called for a redesign of the MAP-A in 2004. The redesigned assessment was intended to meet the needs of students and teachers while complying with the requirements of the federal government.

A general description of the assessment development and standard-setting processes for MAP-A mathematics, communication arts, and science assessments follows. For more detailed information about the assessment development, please refer to Appendix A, Mathematics and

Communication Arts Assessment Development Process, and Appendix B, Science Pilot Assessment Development Process.

Mathematics and Communication Arts

The MAP-A was developed as a collaborative project by Measured Progress, the Assessment Resource Center (ARC) and DESE divisions of Curriculum and Assessment and Special Education. Mathematics and communication arts development began in the 2004-2005 academic year with the discussions of the MAP-A Advisory Committee, made up of stakeholders that included parents, teachers, and school administrators. In addition to this committee, the contractor and DESE called together groups of Missouri educators several times to participate in the development and review process. Special education and general education teachers made up the review groups that developed the AGLEs, in cooperation with DESE and Measured Progress assessment and content specialists. They used the Missouri Show-Me Standards and the Grade Level Expectations (GLEs) to draft and revise AGLEs, which were in turn the basis for the APIs used for assessment with the MAP-A. Prior to their adoption, the AGLEs and APIs were presented to district personnel for review and comment.

After considering concerns expressed by the MAP-A Advisory Committee, chief among which was the paperwork burden on teachers, DESE and Measured Progress drafted an assessment blueprint and piloted mathematics and communication arts assessments. Missouri's Technical Advisory Committee (TAC) reviewed the blueprint prior to administration of the pilot.

In February 2005, the teachers recruited to pilot mathematics and communication arts were required to attend one of four training sessions delivered at various locations around the state. A total of 164 pilot assessments were administered March-April 2005. Pilot teachers provided feedback to the developers through direct contact and responses to a survey administered to each. The pilot assessments were scored in May 2005 at ARC. Measured Progress led table leader training. Sessions were attended by ARC staff and DESE staff. Scorers were asked to provide feedback through a survey administered following the training and scoring.

DESE considered the feedback and suggestions provided by pilot teachers and scorers, along with the input of its advisory groups to make refinements to the MAP-A prior to its initial operational assessment year, 2005-2006. Clarifications were made to training materials and the development of additional samples for teachers was planned. The most significant change, however, was made to the blueprint. In response to serious concerns from teachers about the workload and ability to assess the nine strands in each content area, the number of strands required for assessment at each grade span was decreased from nine to four.

Following the initial operational administration, Measured Progress conducted a standard-setting meeting in Columbia in June 2006 to set cut scores that would be used to determine achievement levels for mathematics and communication arts. Eighty-three panelists, divided into six grade-span and content-area groups, participated in the three-day meeting. Measured Progress employed the modified Body of Work Method, in which panelists are presented with a set of actual student work and are asked to determine which performance level best matches the skills and abilities evidenced in the student work sample.

Individual participants were recruited by Measured Progress and ARC with the goal of empanelling a demographically diverse group that represented a mix of parents, special education teachers, communication arts and mathematics content teachers, and school

administrators. DESE exercised final approval over panelist selection. At the beginning of the meeting, all panelists attended a large-group training containing an overview of the MAP-A, participation criteria, administration information, scoring procedures, overview of the standard-setting process and related issues, and finally specific training about the tasks required of panelists. Following this training, the large group broke into grade-level panels which were led through their tasks over the three-day meeting by a trained facilitator from Measured Progress.

The standard-setting process included three rounds of panelist review. The first consisted of achievement level descriptors review and discussion, review of assessment submissions, and individual cut-point recommendation. The second and third rounds consisted of individual cut-point recommendation after extensive group discussion. Within each round, the panelists first made the middle (Basic-Proficient) cut, then sorted the below Proficient group into Below Basic and Basic, and finally sorted the second group by determining an upper (Proficient-Advanced) cut. Following the second round, the percentage distribution of achievement level impact data was presented to the groups by Measured Progress's psychometrician, to assist them in their round 3 discussions. After the final round, panelists again turned their attention to the achievement level descriptors, and made recommendations for clarifications to the language.

At the conclusion of the meeting, the changes and cut scores recommended by the panelists were reviewed by Measured Progress and DESE. Measured Progress applied smoothing methods and recommended achievement level descriptors and cut-score tables to DESE for consideration by the Missouri State Board of Education. The achievement level descriptors and cut scores were approved by the board and used to generate reports and accountability information for the 2005-2006 school year.

Detailed information about the standard-setting process may be found in the June 2006 MAP-A Standard Setting Report at the DESE website, http://dese.mo.gov/college-career-readiness/assessment/assessment-technical-support-materials.

Science

The development of the science assessment began in the 2006-2007 school year. In addition to the MAP-A Advisory Committee, a Science Assessment Development and Review Committee, also made up of stakeholders that included parents, teachers, and school administrators, provided input to the development process. The AGLE/API development process followed much the same format as that used for the mathematics and communication arts AGLEs and APIs, as did the rest of the development process, including review and comment from groups of Missouri educators, the MAP-A Advisory Committee, and the TAC.

Pilot teacher training for 135 volunteer teachers was conducted in December 2006 at four locations in Missouri. The science pilot was administered to 92 students during the January-March 2007 window, and scored in Columbia in June 2007. As with the other two subjects, surveys were administered to pilot participants, both teachers and scorers, and their responses were considered, along with any face-to-face feedback they provided. The two ideas that emerged involved the provision of information to teachers about administering MAP-A science for two primary reasons: 1) differences in assessment requirements, and 2) teachers' concerns about their own expertise with science content. DESE and Measured Progress made plans to

address these concerns, adding additional information to training materials, providing pathways to science content specialists and planning the expansion of science samples.

Measured Progress, as it did for mathematics and communication arts, used the modified Body of Work method in the standard-setting process for science. The standard-setting meeting took place over two days in the late spring of 2008, following the first operational administration of MAP-A science assessments and followed much the same format as the June 2006 standard-setting meeting. One difference of note in the outcome of the science standard-setting is the establishment of a uniform set of cut scores across all three grade levels in science.

The MAP-A science achievement level descriptors and cut scores were approved by the Missouri State Board of Education and used to generate score reports and accountability data for the 2007-2008 school year. More information about the standard-setting process, and the science standard-setting meeting itself, may be found the DESE website, http://dese.mo.gov/college-career-readiness/assessment/assessment-technical-support-materials.

The initial MAP-A science blueprint differed from that of mathematics and communication arts. It required only two entries, but each contained an activity that addressed two APIs from two different strands. In this way, the science assessment entries paired standards from grade-level-specific science content strands and all-grade-level science process strands. In all, MAP-A science required the assessment of four strands.

NCLB requires technical documentation for all components of Missouri's statewide assessment system, including MAP-A, to be submitted to the United States Department of Education's Office of Elementary and Secondary Education for Peer Review. Following review of a report completed in December, 2009 by Human Resources Research Organization (HumRRO) of the alignment of the MAP-A Science assessment to Missouri's Show-Me Standards and the Science AGLEs, the Peer Review committee assigned to Missouri requested that the state submit a plan and timeline to address the recommendations from the report. One of these recommendations was for the state to review the Science AGLEs for grade appropriateness and accessibility.

As a result, DESE brought together a statewide committee of Missouri practitioners which included administrators of special education, general education science teachers, and special education teachers representing a wide range of grade spans and certification status. The committee spent seven days during the months of March and April 2011 reviewing the Science AGLEs for grade appropriateness and accessibility. At the conclusion of its work, the committee submitted a revised version of the Science AGLEs. After DESE review, the AGLEs were approved and the updated Alternate Performance Indicators were implemented in the 2011-2012 MAP-A testing window administration. Along with the revision of the Science AGLEs, the science blueprint was amended to include four entries, each assessing one API from one of six strands. DESE conducted a standard-setting study following the 2012 MAP-A science administration in the summer of 2012 as well as an alignment study on the MAP-A science assessment in the fall of 2012.

Detailed information about the standard-setting and alignment processes may be found at the DESE website, http://dese.mo.gov/college-career-readiness/assessment/assessment-technical-support-materials .

MAP-A Chronology

Major milestones in the MAP-A development process and subsequent administration of the MAP-A are listed in the chronology below.

1999 – 2000

- MAP-A mathematics and communication arts assessments are administered as voluntary assessments.

2000 – 2003

- MAP-A mathematics and communication arts assessments are required and administered to eligible students at ages 9, 13, and 17.

2004 – 2005

- MAP-A mathematics and communication arts assessments are administered to eligible students in grades 4, 8, and 11.
- DESE contracts with Measured Progress for development of a redesigned MAP-A to assess mathematics and communication arts.
- Development involves multiple groups of stakeholders and advisors.
- Mathematics and communication arts assessments are piloted.

2005 – 2006

- Revisions based on stakeholder feedback are made to MAP-A design.
- Operational assessment in mathematics and communication arts commences.
- MAP-A mathematics assessments are administered to eligible students in grades 3 through 8 and 10; communications arts assessments are administered in grades 3 through 8 and 11.
- Standard setting for mathematics and communication arts is conducted and the resulting cut scores are approved by the Missouri State Board of Education.
- DESE contracts with Measured Progress for development of MAP-A science assessment. Development involves multiple groups of stakeholders and advisors.

2006 – 2007

- Revisions in response to stakeholder feedback are made to MAP-A.
- Mathematics and communication arts are administered to eligible students in grades 3-8 and one grade in high school for the second year.
- The MAP-A science component is developed and piloted; Measured Progress documented the science development process. This documentation may be found in Appendix B.

2007 – 2008

- Revisions in response to stakeholder feedback are made to MAP-A.
- Mathematics and communication arts are assessed with MAP-A for the third year.
- The MAP-A science component becomes operational and is assessed at grades 5, 8, and 11.
- Measured Progress conducts standard-setting meeting for the science assessment and the resulting cut scores are approved by the Missouri State Board of Education.

2008 – 2009

- Updates and revisions in response to stakeholder feedback are made to MAP-A training materials and resources.
- Mathematics and communication arts are assessed with MAP-A for the fourth year; science is assessed with the MAP-A for the second year.
- DESE offers MAP-A scoring training to teachers administering the MAP-A as professional development.

2009 – 2010

- Updates and revisions in response to stakeholder feedback are made to MAP-A training materials and resources.
- Mathematics and communication arts are assessed with MAP-A for the fifth year; science is assessed with the MAP-A for the third year.
- Supplemental professional development is offered through Regional Professional Development Centers to teachers in the form of MAP-A scoring training.
- Science alignment study is conducted by HumRRO

2010-2011

- Updates and revisions in response to stakeholder feedback are made to MAP-A training materials and resources.
- Mathematics and communication arts are assessed with MAP-A for the sixth year; science is assessed with the MAP-A for the fourth year.
- Science AGLE revision is conducted by DESE.

2011-2012

- Updates and revisions in response to stakeholder feedback are made to MAP-A training materials and resources.
- Mathematics and communication arts are assessed with MAP-A for the seventh year; science is assessed with the MAP-A for the fifth year.
- Science AGLE revision is approved by DESE.
- Amended science blueprint is implemented.
- Pearson conducts standard-setting meeting for the science assessment and the resulting cut scores are approved by the Missouri State Board of Education.
- Science alignment study is conducted by HumRRO.

2012-2013

- Updates and revisions in response to stakeholder feedback are made to MAP-A training materials and resources.
- Mathematics and communication arts are assessed with MAP-A for the eighth year; science is assessed with the MAP-A for the sixth year.

2013-2014

- Updates and revisions in response to stakeholder feedback are made to MAP-A training materials and resources.
- Mathematics and communication arts are assessed with MAP-A for the ninth year; science is assessed with the MAP-A for the seventh year.

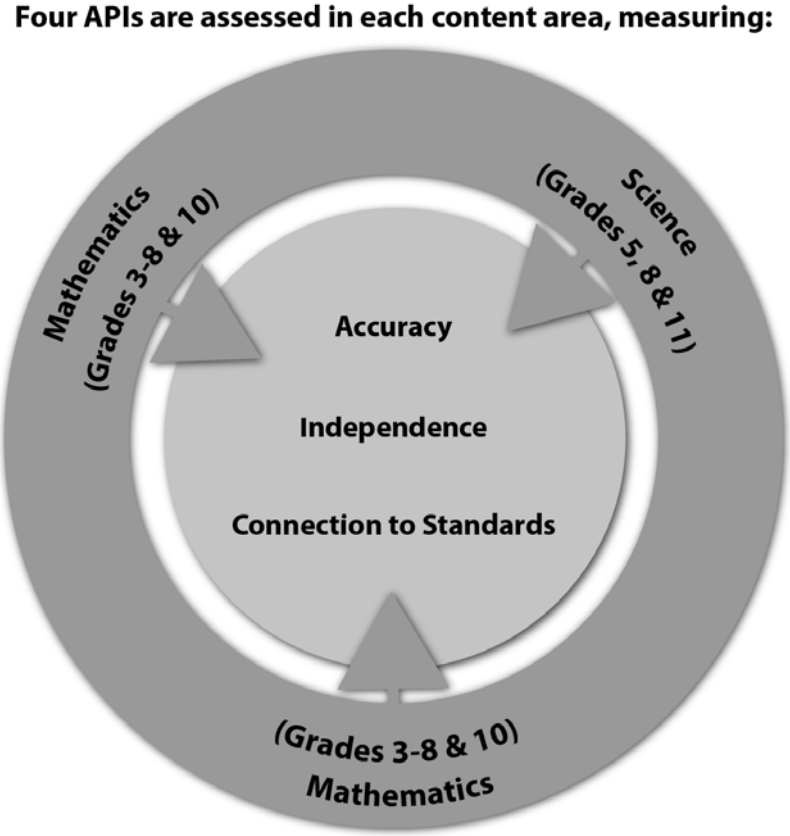
2014-2015

- Updates and revisions in response to stakeholder feedback are made to MAP-A training materials and resources.
- Mathematics and communication arts are not assessed with MAP-A; science is assessed with the MAP-A for the eighth year.

Introduction to the MAP-A Process

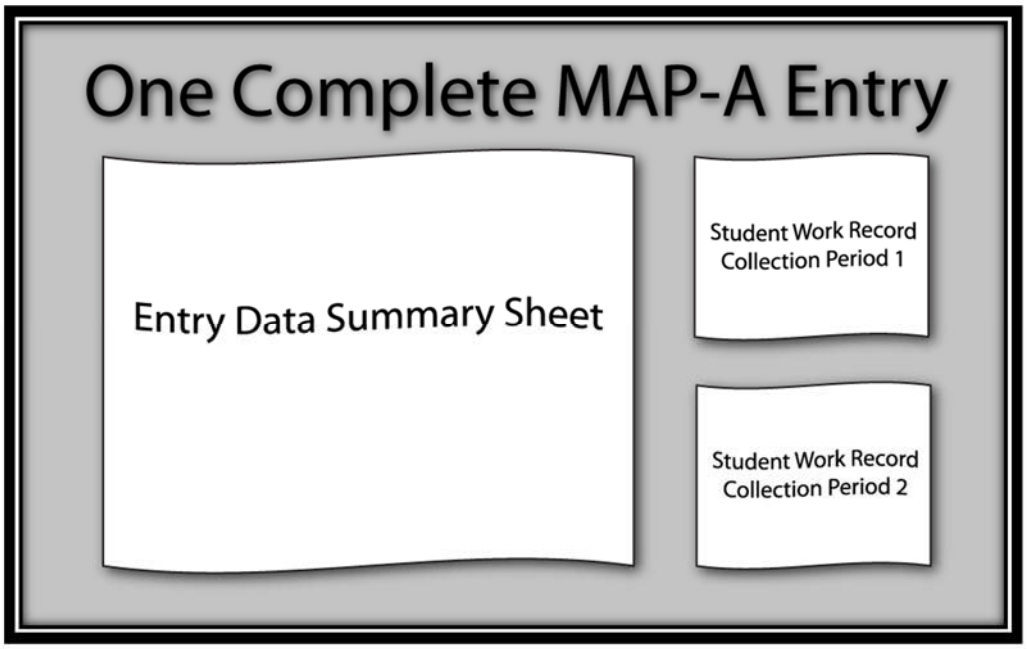
The MAP-A calls for information about the performance of students with significant cognitive disabilities on assessment activities designed and implemented by their teachers. The assessment activities are designed to provide evidence of student knowledge and ability in mathematics, communication arts, and science. The MAP-A assesses accuracy, independence, and connection to the standards on four APIs in each subject.

Figure 1. MAP-A Assessment Design



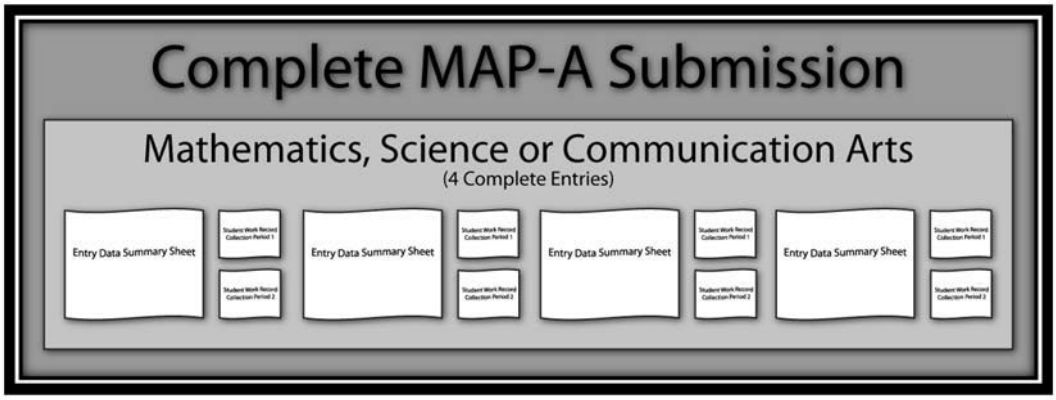
Teachers design activities to assess these APIs; they are trained to build their activities to align with the standards to assess and the student's highest academic functioning level. Activity descriptions for each API are submitted in Student Work Record forms in the student's binder. Teachers record data for an API three times during each of two collection periods, altogether producing six data points and two Student Work Records for that entry. These data points are averaged together on an Entry Data Summary Sheet to create that entry's Accuracy and Independence percentages.

Figure 2. MAP-A Entry



Each complete MAP-A mathematics, communication arts, and science submission contains four entries (one for each API).

Figure 3. MAP-A Submission



All submissions for a student's MAP-A are combined in that student's binder along with a Table of Contents Checklist and Validation Form. Completed binders are returned to ARC for processing and scoring.

Scorers review submitted binders and assign rubric scores to each entry. These scores correspond to student Level of Accuracy and Level of Independence averages provided by teachers. A Connection to the Standards rubric score is determined by considering whether the assessment activity connects to the API and if the activity demonstrates application of the skill in the API. When scoring irregularities occur (e.g., no connection to the API, missing documentation), scorers record the appropriate comment codes as well as the rubric score. Final entry rubric scores are added together to create the raw score for each content area. DESE-approved cut scores are

used to assign achievement levels for each assessment.

Table 1. Condensed MAP-A Rubric

Rubric	Score-Point 4	3	2	1	No Score
Level of Accuracy	76-100%	51-75%	26-50%	0-25%	Entry contains insufficient evidence to score.
Level of Independence	76-100%	51-75%	26-50%	0-25%	Entry contains insufficient evidence to score.
Connection to the Standards		Entry contains evidence of applying the API in two standards-based activities, one per collection period.	Entry contains evidence of applying the API in one standards-based activity, one out of two collection periods.	Entry contains some evidence of a connection to the API.	Entry contains insufficient evidence of connection to the API.

Teachers and individuals familiar with MAP-A administration and evaluation routinely use many acronyms and terms that may be unfamiliar to all readers. Several common terms are outlined below.

Table 2. Common MAP-A Terms

Term	Definition
Acquisition	Activities that demonstrate acquisition focus on practicing skills rather than applying them for a purpose.
AGLE	Alternate Grade Level Expectations
API	Alternate Performance Indicators
Application	Activities that demonstrate application require the student to apply skills for purposes other than practicing.
CTS	Connection to the Standards
Entry	A student binder component that includes an Entry/Data Summary Sheet, two Student Work Records, and optional Student Work samples.
IEP	Individualized Education Program
Validation Form	A student binder component that includes the student's mode of communication, the names of individuals who reviewed and/or contributed to the development or administration of the student's MAP-A, and the signature of the administrator who approved the binder for final submission.
Work Record	An entry component that contains the Task/Activity, Level of Accuracy, and Level of Independence descriptions.

Operational Assessment Administration

The MAP-A was administered in the spring of 2015 to students meeting the Missouri's alternate assessment eligibility criteria. Science assessments were administered to students in grades 5, 8, and 11. Students from 437 districts participated in the MAP-A; 2,615 students participated in science.

Eligible Students

All students are required to participate in the Missouri Assessment Program in one of four ways: 1) MAP Grade-Level assessments, 2) MAP End-of-Course assessments, 3) MAP Grade-Level or End-of-Course assessments with accommodations, or 4) the MAP-A.

The decision as to how a student with disabilities will participate in the state's accountability system is made by the student's Individualized Education Program (IEP) team using DESE-established criteria. If the IEP team for a student with a disability answers "yes" to all five of the following eligibility questions, then the student is eligible for MAP-A participation.

MAP-A Participation Eligibility Criteria

Yes	No	
__	__	1. The student has a demonstrated significant cognitive disability and adaptive behavioral skills. Therefore, the student has difficulty acquiring new skills, and skills must be taught in very small steps.
__	__	2. The student does not keep pace with peers, even with the majority of students in special education, with respect to the total number of skills acquired.
__	__	3. The student's educational program centers on the application of essential skills to the Missouri Show-Me Standards.
__	__	4. The IEP team, as documented in the IEP, does not recommend participation in the MAP assessments (Grade-Level or End-of-Course) or taking the MAP with accommodations.
__	__	5. The student's inability to participate in the MAP Grade-Level or End-of-Course assessments is not primarily the result of excessive absences; visual or auditory disabilities; or social, cultural, language, or economic differences.

In an attempt to provide more information for educators charged with making the MAP-A eligibility decision, DESE provided statements as a supplement to criterion #3. These statements may be used by IEP teams in identifying students whose educational programs center on the **application of essential skills** to the Missouri Show-Me Standards:

1. The student's reading ability is limited and, as such, the student acquires information primarily through other methods.
2. The student's ability to demonstrate knowledge by writing or speaking is limited; thus, the student must often use other methods to express ideas and share information.
3. The student requires significant supports to access the general education curriculum while demonstrating modest progress in that curriculum.

4. The student typically has difficulty solving novel problems or using newly acquired skills in differing situations.
5. The student's educational priorities primarily address essential skills that will be used in adult daily living.
6. The student's post-secondary outcomes will likely require supported or assisted living.
7. The student requires instruction in small groups or on a one-to-one basis, with frequent prompts and guidance from adults.

The grade-level MAP and End-of-Course assessments provide access to the vast majority of students. Therefore, approximately 1% of Missouri students assessed are expected to participate in the MAP-A. In accordance with NCLB regulation 34 CFR 200.13 Adequate Yearly Progress in General, if necessary Missouri would apply a 1% cap to the number of proficient and advanced scores based on the MAP-A that may be included in AYP calculations at both the state and district levels.

District test coordinators were required to enroll MAP-A eligible students in the MAP-A through ARC in fall 2014. This triggered delivery of a set of student-specific materials to the districts for each student enrolled in the MAP-A and an expectation that a MAP-A would be submitted for scoring for that student in spring 2015.

Assessment Blueprint/Design

The MAP-A is a performance-based assessment that promotes enhanced capacities and integrated life opportunities for students with severe disabilities. One key purpose is to capture evidence of student learning. Another key purpose, in accord with high-quality assessment practices, is to provide information upon which to base ongoing development of curricula and instruction that are responsive to individual student needs. Students with significant cognitive disabilities are valued and contributing members of their school and community. Missouri implements and continues to improve the MAP-A to meet the needs of students and teachers as well as to comply with the requirements of the federal government.

The MAP-A consists of a portfolio of data and supporting evidence collected by an instructional team. It provides information on a student's knowledge and skills in communication arts, mathematics, and science. The MAP-A assesses accuracy, independence, and connection to the standards on two APIs in each of two strands in communication arts and mathematics; the MAP-A also assesses four APIs selected from six strands in science. Tables 3, 4, and 5 contain the assessment blueprints for the three subjects.

Table 3. Assessment Blueprint for Mathematics *(not assessed 2014-2015)*

Content Area	Grade Focus	Title of Strand
Mathematics	Required for Grades 3-8 and 10	Numbers and Operations **(NO)**
	Required for Elementary Grades 3, 4, and 5	Algebraic Relationships **(AR)** *and/or* Geometric and Spatial Relationships **(GS)**
	Required for Middle School Grades 6, 7, and 8	Data and Probability **(DP)**
	Required for High School Grade 10	Measurement **(ME)**

Table 4. Assessment Blueprint for Communication Arts *(not assessed 2014-2015)*

Content Area	Grade Focus	Title of Strand
Communication Arts	Required for Grades 3-8 and 11	Reading: Develop and apply skills and strategies to the reading process (**RD** *and/or* **RP)**
	Required for Elementary Grades 3, 4, and 5	Writing: Compose well-developed text using standard English conventions **(WC)**
	Required for Middle School and High School Grades 6, 7, 8, and 11	Writing: Apply a writing process in composing text or write effectively in various forms and types of writing **(WP)**

Table 5. Assessment Blueprint for Science

Content Area	Grade Focus	Title of Strand
Science	Required for Elementary School Grade 5	• Strand 5: Processes and Interactions of the Earth's Systems **(ES)**
		• Strand 6: Composition and Structure of the Universe and the Motion of the Objects within it **(UN)**
		• Strand 7: Scientific Inquiry (IN) **or** Strand 8: Impact of Science, Technology, and Human Activity **(ST)**
		• Strand 3: Characteristics and Interactions of Living Organisms **(LO)** **or** Strand 4: Changes in Ecosystems and Interactions of Organisms with Their Environment **(EC)**
	Required for Middle School Grade 8	• Strand 1: Properties and Principles of Matter and Energy **(ME)**
		• Strand 2: Properties and Principles of Force and Motion **(FM)**
		• Strand 7: Scientific Inquiry **(IN)** **or** Strand 8: Impact of Science, Technology, and Human Activity **(ST)**
		• Strand 5: Processes and Interactions of the Earth's Systems **(ES)** **or** Strand 6: Composition and Structure of the Universe and the Motion of the Objects within It **(UN)**
	Required for High School Grade 11	• Strand 3: Characteristics and Interactions of Living Organisms **(LO)**
		• Strand 4: Changes in Ecosystems and interactions of Organisms with Their Environment **(EC)**
		• Strand 7: Scientific Inquiry **(IN)** **or** Strand 8: Impact of Science, Technology, and Human Activity **(ST)**
		• Strand 1: Properties and Principals of Matter and Energy **(ME)** **or** Strand 2: Properties and Principals of Force and Motion **(FM)**

Mathematics and communication arts are assessed at grades 3 through 8. Mathematics is also assessed at grade 10. Communication arts is also assessed at grade 11. Science is assessed at grades 5, 8, and 11. All three content areas require assessment of four different APIs. APIs for MAP-A entries must be selected from particular strands within each content area, depending upon the student's grade level.

For example, the mathematics Measurement strand (ME) includes 55 APIs, from which two must be selected for a 10th-grade student's MAP-A mathematics assessment, along with two APIs from the Numbers and Operations strand (NO). The following is a sample of nine APIs from the Measurement strand.

Alternate Performance Indicators (APIs)

Justify and use the appropriate unit of measure (linear, time, weight).

ME1.1. Recognize, compare, and order attributes such as length and weight.

a. Compare and communicate the length of 2 objects directly, using words such as "bigger," "smaller," "longer," "shorter," and "taller."

b. Compare and communicate the weight of 2 objects directly, using words such as "heavier," and "lighter."

c. Engage in experiences to connect number with length, using both conventional rulers and manipulative units that are standard units, such as centimeter cubes.

d. Engage in experiences to connect number with weight, using balance and spring scales.

e. Select and identify the appropriate tool for the attribute being measured.

f. Show understanding of unit iteration for length measurement (e.g., placing units end to end in some manner, with no gaps).

g. Use repetition of a single unit to measure something larger than the unit (e.g., measuring the length of the room with a single meter stick).

h. Use appropriate unit for the attribute being measured.

Complete API lists may be found in the *Instructor's Guide and Implementation Manual* and/or at DESE's MAP-A web page.[1]

Once the APIs are selected, the MAP-A requires that data for each API be collected over two collection periods to form a MAP-A entry. For each entry, three data points per collection period must be recorded on the Entry/Data Summary Sheet. One of these three data points per collection period must be further described and documented on a Student Work Record. Actual student work, appropriate for inclusion in the portfolio, is submitted with the student work record.

A **complete MAP-A entry** is defined, at a minimum, as one Entry/Data Summary Sheet and two Student Work records documenting six data points for each API. Each subject requires submission of four entries. Because there are four APIs, and four entries required, a student's content area submission will contain documentation for 24 data points at a minimum. In all, there is a total of 72 MAP-A data points per student participating in mathematics, communication arts, and science. Table 6 below outlines the requirements.

[1]http://dese.mo.gov/college-career-readiness/assessment/map-a#Manuals

Table 6. Mathematics, Communication Arts *(not assessed 2014-2015)* **and Science Data Collection and Submission Requirements**

Entry	APIs per Entry	Collection Period	Data Collection Required	Forms Required		Min. Total # of Pages
1	1	1	3 data points	1 Entry/Data Summary Sheet	2 Student Work Records	12
		2	3 data points			
2	1	1	3 data points	1 Entry/Data Summary Sheet	2 Student Work Records	
		2	3 data points			
3	1	1	3 data points	1 Entry/Data Summary Sheet	2 Student Work Records	
		2	3 data points			
4	1	1	3 data points	1 Entry/Data Summary Sheet	2 Student Work Records	
		2	3 data points			

Steps for MAP-A Administration

The administration process follows twelve steps that take the teacher from determining student eligibility to the point of submitting the assessment. These steps are outlined in the *Instructor's Guide and Implementation Manual* provided to teachers. That manual provides detailed information on what evidence to collect and how to do so for each student and also provides many samples for teachers to refer to during the process. The twelve steps are as follows:

A Twelve-Step Procedure for Completing the MAP-A

1. **Verify student eligibility for participation in the MAP-A. Refer to the student's IEP.**
   For information about eligibility see the Participation Eligibility Criteria established by DESE.

2. **Determine the composition of the instructional team that will assess the student and fully inform all participants about the MAP-A.**
   The instructional team may include teachers, administrators, physical therapists, speech therapists, occupational therapists, paraprofessionals, job coaches, parents or guardians, and the student, when appropriate. **The student's case manager/teacher is responsible for the coordination of the assessment**. The case manager/teacher should fully inform all participants on the instructional team about the alternate assessment. Other professionals responsible for assisting the case manager/teacher in collecting information about the student should be aware of the MAP-A requirements and their roles in administering the MAP-A. Members of the instructional team are listed on the MAP-A validation form. The instructional team may have members in common with the IEP team, but they are NOT the same group.

3. **Identify the mandatory strands in each content area.**
   The instructional team should refer to the Assessment Blueprint prior to beginning collection of evidence for the MAP-A.

4. **Select Alternate Performance Indicators (APIs) for each required content-area strand.**
   The instructional team should refer to the Alternate Performance Indicators for a list of appropriate grade-level APIs for each strand.

   - For mathematics and communication arts, **two APIs per strand** are required.
   - For science, **one API per grade-appropriate strand** is required.

5. **Review the requirements for documentation for the MAP-A.**
   The following forms are required to complete documentation for each API:

   - **Form 1:** Entry/Data Summary Sheet
     This form is used to determine student scores for the rubric dimensions *Level of Accuracy* and *Level of Independence*. The following are included on the Entry/Data Summary Sheet:
     - Student identification
     - Content area and strand identification
     - API identification and description
     - Summary data chart
   - **Form 2:** Student Work Record
     This form is used to determine the student's score for the rubric dimension *Connection to the Standards*. In order to obtain full credit for this rubric

dimension, the Student Work Record must show *application* of the API in standards-based activities. The following are included on the Student Work Record:

- Student identification
- Content area and strand identification
- API identification and description
- Activity description
- Description and evaluation of student performance

6. **Determine the data collection system for documentation of student performance.**
   The instructional team selects the APIs and determines how student performance will be documented. The team should ask the following questions when planning for data collection:
   - How was the activity designed?
   - What type of data will be collected?
     - Discrete trials
     - Task analyses
     - Time intervals
     - Accuracy rates
   - How will the data be collected and organized?
   - Who will collect the data?
   - When will the data be collected?
   - How will data be converted into percentage scores?

7. **Collect and record data throughout the assessment period.**
   There are two required collection periods for the recording of data on the Entry/Data Summary Sheet. Only data collected during the identified collection periods should be included on the data sheets. There must be three data points per collection period, one of which is linked to a Student Work Record.

8. **Select a Student Work Record to include in the MAP-A for each collection period.**
   The data from the Student Work Records submitted must be documented on the Entry/Data Summary Sheet. Make sure the activity shows evidence of application of the API.

9. **Complete the Student Work Record.**

10. **Complete the Entry/Data Summary Sheet for each assessed API.**
    There are two steps to completing the Entry/Data Summary Sheet prior to submission of the MAP-A:
    - Determine API percentage averages.
      - **a.** Average the two scores for *Level of Accuracy*.
      - **b.** Average the two scores for *Level of Independence*.
    - Indicate the Student Work Record included for each collection period of the API.

11. **Assemble the MAP-A documentation.**
    Once all of the required documentation has been completed, the teacher should assemble the MAP-A as directed in the Table of Contents Checklist.

12. **Submit completed MAP-A.**
    Submit completed MAP-A to your district test coordinator on or before the MAP-A return

Implementation Schedule

The schedule for the MAP-A began with the September 2014, administration training and continued with trainings conducted by RPDC staff beginning in September 2014. Assessment materials were shipped to districts December 2014 through early January 2015, and two distinct data collection periods spanned January through late February 2015. MAP-A submissions were returned to ARC in March 2015 for scoring. Table 8 outlines this timeline.

Table 8. 2015 MAP-A Timeline

Event	Dates
Enrollment Window	September 8 – November 3, 2014
Transfer Administration Date	January 2, 2015
Collection Period 1	January 5 – January 30, 2015
Collection Period 2	February 2 – February 27, 2015
Return Deadline	March 6, 2014

Participation

MAP-A participation totaled 2615 students. A summary of Missouri student participation in the 2015 MAP-A assessment is provided in Table 9. See the Scoring and Reporting section for additional information regarding student participation and performance.

Table 9. 2015 MAP-A Participation

Content Area	Grade Span/Level	Students Participating
Science	5	897
	8	890
	11	828

Scoring and Reporting

MAP-A scoring was conducted at the Assessment Resource Center (ARC). Scoring took place over several weeks beginning in April and continuing through June 2014.

Scoring Rubric

The scoring rubric is the basis for determining the student scores on the MAP-A. Three dimensions are scored:

1. Level of accuracy. This dimension reflects how well the student understands the concept(s) being assessed.
2. Level of independence. This dimension reflects the extent to which the student is able to perform without assistance from the examiner.
3. Connection to the standards. This dimension reflects whether the assessment is clearly linked to the Show-Me Standards.

Scorers review the entries submitted and assign rubric scores for each of the three dimensions. Level of accuracy and level of independence are scored using a four-point rubric. Connection to the standards is scored using a three-point rubric. The total entry score is a simple sum of these three, and ranges from 0 to 11 points. A sum of the entry scores for the four entries required for mathematics, communication arts, and science makes up the total raw score for that subject area. The total raw score ranges from 0 to 44 points.

Table 10 shows the rubric dimensions.

Table 10. MAP-A Rubric

Rubric	Score Points				
	4	3	2	1	No Score
Level of Accuracy	Student performance of skills "based on Alternate Performance Indicators" demonstrates a **high level of understanding** of concepts. **76–100% Accuracy**	Student performance of skills "based on Alternate Performance Indicators" demonstrates **some understanding** of concepts. **51–75% Accuracy**	Student performance of skills "based on Alternate Performance Indicators" demonstrates a **limited understanding** of concepts. **26–50% Accuracy**	Student performance of skills "based on Alternate Performance Indicators" demonstrates a **minimal understanding** of concepts. **0–25% Accuracy**	Entry contains insufficient information to determine a score.
Level of Independence	Student requires **minimal** verbal, visual, and/or physical assistance to demonstrate skills and concepts. **76–100% Independence**	Student requires **some** verbal, visual, and/or physical assistance to demonstrate skills and concepts. **51–75% Independence**	Student requires **frequent** verbal, visual, and/or physical assistance to demonstrate skills and concepts. **26–50% Independence**	Student requires **extensive** verbal, visual, and/or physical assistance to demonstrate skills and concepts. **0–25% Independence**	Entry contains insufficient information to determine a score.
Connection to the Standards	--	There is evidence of applying the Alternate Performance Indicator in **two** standards-based activities, one per collection period.	There is evidence of applying the Alternate Performance Indicator in at least **one** standards-based activity, **one out of two collection periods**.	There is some evidence of a connection to the Alternate Performance Indicator.	There is insufficient evidence of a connection to the Alternate Performance Indicator.

MAP-A data submissions are not always complete and may not follow submission guidelines. Table 11 shows potential data irregularities, the rules used to address them, and the frequencies at which these irregularities appeared in the MAP-A entries for 2015.

Table 11. Scoring Rules

Code	Data Irregularity	Scoring Rule	# of Appearances in Scored 2015 Entries	% of Total Scored 2015 Entries
01	No dates given on Entry/Data Summary Sheet and on Student Work Records.	Entry is assigned a "No Score" for each dimension of the rubric.	1	0.01
02	Missing Entry/Data Summary Sheet.	Entry is assigned a "No Score" for each dimension of the rubric.	7	0.07
03	A collection period does not have a minimum of three data points.	Entry is assigned a "No Score" for each dimension of the rubric.	138	1.32
04	An entry does not include at least one Student Work Record per Collection Period.	Entry is assigned a "No Score" for each dimension of the rubric.	37	0.35
05	A submitted Student Work Record for an entry does not connect to the API/s.	Entry is assigned a "No Score" for each dimension of the rubric.	920	8.8
06	One out of two collection periods is incomplete.	Entry is assigned a "No Score" for each dimension on the rubric.	0	0.00
07	No API/s identified on a Student Work Record or Entry Data/Summary Sheet.	The collection period is considered incomplete. Entry is assigned a "No Score" for each dimension on the rubric.	0	0.00
08	The API/s is/are not grade-span appropriate.	The collection period is considered incomplete. Entry is assigned a "No Score" for each dimension on the rubric.	0	0.00
09	A single API is used in more than one entry.	The first instance is scored. In the second instance, the entry is assigned "0 Data Points" in both collection periods and "No Score" for each dimension of the rubric.	0	0.00

Table 11. Scoring Rules (contd.)

Code	Data Irregularity	Scoring Rule	# of Appearances in Scored 2015 Entries	% of Total Scored 2015 Entries
11	Missing entry.	Entry is assigned "0 Data Points" in both collection periods and "No Score" for each dimension on the rubric.	153	1.46
12	API/s is/are not consistent across the 2 collection periods.	Entry is assigned a "No Score" for each dimension of the rubric.	0	0.00
13	Dates on the Entry/Data Summary Sheet and Student Work Records are not within the timeframes of the collection periods.	Any data from dates outside of the timeframes is not used for scoring.	0	0.00
14	One or more Student Work Records shows acquisition rather than application of the API/s.	The activity in these collection periods cannot be considered application.	1870	17.88
15	Student work sample or piece of tangible student work submitted without a Student Work Record attached.	The activity in this collection period cannot be considered application.	0	0.00
16	Student Work Record missing task/activity description.	The activity in this collection period cannot be considered application.	1	<0.01
17	Submitted percentages are miscalculated.	Scorer corrects percentages.	183	1.75
18	Percentage calculations for Accuracy or Independence cannot be verified for a Student Work Record.	Percentage for Accuracy or Independence for the Student Work Record is replaced with zero and entry average is recalculated to determine rubric score.	509	4.87

More information regarding scoring criteria may be found in Appendix G.

Scorer Selection

ARC has many years' experience hiring and training scorers to read, evaluate, and score open-ended assessments (fill-in-the-blank, short answer, short or long essay, and portfolio) for students at the primary, secondary, and post-secondary educational levels in subject areas including reading/language arts, mathematics, science, and social studies. Emphasis is placed on the maintenance of security and confidentiality of tests at all times. Scorers consult with scoring facilitators about scoring questionable responses to determine how to score them and attend

regularly scheduled meetings in order to identify and provide input for solving problems or potential problems. Facilitators exercise functional supervision over reader/scorers and/or other staff as necessary.

ARC recruited scorers and facilitators specifically for the MAP-A program. Minimum qualifications for MAP-A scorers include a baccalaureate degree, strong communication skills, and demonstrated ability to critically review printed material. In addition, MAP-A scoring facilitators have prior scoring experience, strong facilitation skills, and the ability to instruct scorers regarding the meaning and application of scoring rubrics. Preferred qualifications for MAP-A scorers include previous experience scoring open-ended assessments, teaching, editing, and/or participating in structured analysis.

Six scorers and one scoring facilitator scored the 2014-2015 MAP-A submissions from March through June 2015. Scorers and the scoring facilitator were required to sign nondisclosure agreements and agreed to maintain the security of MAP-A materials at all times.

Scorer Training

Scorer candidates participated in training sessions led by MAP-A experts that involved paper-and-pencil scoring training. Scorer training focused on the MAP-A rubric and scoring rules. Scorers were given examples of typical student work illustrating various rubric scores and scoring decisions. Examples of "difficult" submissions presenting a variety of scoring challenges were included. Scorer training also included an emphasis on applying the rubric and decision rules as trained, guarding against bias. Following training, scorer candidates were given qualifying tests. If they passed these tests, candidates were certified to score the MAP-A. After they qualified, scorers participated in further hands-on training that consisted of additional MAP-A scoring exercises and the review of MAP-A submissions scored the previous year. See Appendix H for resources used in MAP-A scorer training.

The scoring facilitator began MAP-A training earlier than the remaining scorer candidates. The participation in intensive training sessions and successful completion of qualifying tests were initial activities in the MAP-A scoring window. In addition to these tasks, the scoring facilitator also assisted with screening scorer candidates.

Scoring Procedures

The facilitator functioned as day-to-day monitor of MAP-A scoring, and conducted retraining using materials approved by the ARC MAP-A program staff. The facilitator met with ARC MAP-A program staff on a regular basis to discuss scoring congruence and MAP-A submission irregularities. A blind second read was conducted on a randomly selected set of portfolios, 35% of the 2015 MAP-A submissions.[2] The facilitator conducted resolution reads on portfolios that contained rubric score disagreements between scorers. In these cases, the facilitator's score prevailed as score of record. In addition, highly qualified senior scoring or program staff audited approximately 3% of MAP-A submissions at each grade span and circulated pre-scored submissions during the scoring window. In cases of disagreement with the initial score, the resolution or audit-read score replaced the initial score as the score of record. The facilitators had access to a variety of quality control information, monitored several MAP-A scoring agreement

[2] The initial scoring design called for a read-behind strategy in which the original score is verified and when necessary, corrected by an expert rater. Historically, the MAP-A read-behind rates ranged from 20% to 100%.

reports throughout each scoring day, and used this information to assist, recalibrate, or retrain scorers as necessary. Scorers were required to maintain acceptable agreement rates (an average of 80% across the three rubric dimensions).

To organize the flow of work during a typical day, The MAP-A facilitator outlined the basic tasks and order of work in a simple-to-follow set of instructions.

Steps for Scorers

1. Take one MAP-A binder from the "In Box."
2. Verify that the student name and grade level on the MAP-A binder match the information in the MAP-A scoring interface.
3. Score according to directions.
4. Place completed MAP-A binder in the "Second Read Box," "Resolution Read Box," or "Completed Binder Box."
5. Repeat process as needed.

Steps for Scoring Facilitator

1. Stock the "In Box" with unscored MAP-A binders.
2. Conduct resolution read on MAP-A binders from the "Resolution Read Box."
3. Place validated MAP-A binders in the "Completed Binder Box."
4. Repeat process as needed.

To promote scoring consistency, MAP-A submissions were sorted and scored by grade span to allow scorers and facilitators to focus on one set of APIs for a prolonged period of time. The content strands and APIs assessed with the MAP-A change from grade span to grade span. Following completion of an entire grade span, the facilitators conducted training to calibrate scorers to the next set of APIs.

Reporting

Paper reports were created at the individual student level and at the district level. Two separate student-level reports were created, one for parents/guardians and one for teachers. Paper reports were printed at ARC or at the University of Missouri Printing Services, located in ARC's building. The score data did not leave ARC and the electronic prepress files were returned with the paper products. Paper reports were sent to both the district of residence and the district of attendance for each student as appropriate. A description of the paper reports follows and report samples may be found in Appendix I.

Reports

Individual Student Report–Parent/Guardian and Teacher

This report contained overall achievement level for a single content area, achievement level descriptors, raw rubric scores, and APIs assessed for each of the required entries. The only difference between the student-level reports was that teacher reports included comments related to any submission irregularities in a student's MAP-A so that teachers could learn to make correct submissions in the future.

API History Report

The Individual Student API History Report listed APIs assessed in 2014-2015 and, if information is available, those assessed in previous years. APIs that were assessed with the MAP-A in more than one year are noted. This report is provided for informational purposes and is meant to assist

administrators, teachers, and parents in tracking the breadth and depth of content assessed with the MAP-A from year to year across a student's educational span.

Student Record Label
The label contained assessment year and achievement level information.

District Report
This report summarized data based on student district of residence, and compared district performance by content area, grade span, and achievement level to overall state performance.

State Schools Building Report
This report was similar to the District Report but compared student data from one Missouri Schools for the Severely Disabled (MSSD) building by content area, grade span, and achievement level to overall MSSD performance.

State Schools Report
This report was similar to the District Report but compared student data from one MSSD building by content area, grade span, and achievement level to overall state performance.

State Schools District Report
This report was similar to the District Report but contained a summary of data of students who attend all MSSD buildings and compared overall MSSD performance by content area, grade span, and achievement level to overall state performance.

Report packages sent to districts included the mathematics, communication arts, and science reports for students who were enrolled or assessed in the district.

Reporting Decision Rules

Reports included achievement levels based upon the application of cut scores that may be found in Appendix E. Table 12 outlines the decision rules used for reporting of MAP-A scores.

Table 12. 2015 MAP-A Score Reporting Rules

Achievement Level	
Below Basic	Cut scores applied. *At least one data point recorded in content area submissions.*
Basic	Cut scores applied.
Proficient	Cut scores applied.
Advanced	Cut scores applied.
Level Not Determined	No assessment data points are provided in content-area-required entries.
Participation	
Participating	Enrolled students for whom MAP-A binders are returned for scoring with evidence of at least a partial attempt to collect data.
Non-participating	Enrolled students for whom empty or no MAP-A binders are returned for scoring.
Accountability	
Accountable	All enrolled students, less those who meet health waiver or enrollment exemptions.
Reportable	All accountable students less Level Not Determined and Non-participating students.
Health Waiver	Approved on an individual basis by DESE committee composed of representatives from Special Education; Assessment; and Accountability, Data and Accreditation.
Enrollment Exemptions	Students who moved in or out of the district after January 2, 2015.

Student Performance

The following tables present information regarding 2015 MAP-A student performance and participation.

Table 13. 2015 Students Tested Using MAP-A by Grade Level

Grade Level	MAP-A Students	Total MO Students	% MAP-A
5	897	67,787	1.3%
8	917	67,867	1.3%
11	828	68,188	1.2%
Total	**2615**	**203,842**	**1.3%**

Table 14. 2015 MAP-A Achievement Level Distribution

Grade Span	Achievement Level	Science	
		n	%
Grades 5,8,11	Level Not Determined	39	1.49
	Below Basic	71	2.72
	Basic	182	6.96
	Proficient	639	24.44
	Advanced	1684	64.4
	Prof & Adv	2323	88.83
Grade 5	Level Not Determined	14	1.56
	Below Basic	38	4.24
	Basic	55	6.19
	Proficient	198	22.07
	Advanced	592	66
	Prof & Adv	790	88.07
Grade 8	Level Not Determined	11	1.24
	Below Basic	19	2.13
	Basic	81	9.1
	Proficient	241	27.08
	Advanced	538	60.45
	Prof & Adv	779	87.52
Grade 11	Level Not Determined	14	1.69
	Below Basic	14	1.69
	Basic	46	5.56
	Proficient	200	24.15
	Advanced	554	66.91
	Prof & Adv	54	91.06

Table 17. 2015 MAP-A Science Achievement Level Distribution by Grade Level

Grade	Total Students	Level Not Determined & Below Basic *		Basic		Proficient		Advanced		Prof & Adv	
		n	%	*n*	%	*n*	%	*n*	%	*n*	%
5	897	52	5.79	55	6.13	198	25.82	592	66	790	88.07
8	890	30	3.37	81	9.1	241	27.08	538	60.45	779	87.52
11	828	28	3.38	46	5.56	200	24.15	554	66.91	754	91.06
Total	2615	110	4.20	182	6.96	639	24.44	1684	64.4	2323	88.83

* Level Not Determined and Below Basic data combined due to small sample size.

Table 20. 2015 MAP-A Science Achievement level Distribution by Gender, Ethnicity, Primary Disability, Student Status, ELL Status, and Classroom Instruction

	Level Not Determined		Below Basic		Basic		Proficient		Advanced		Prof & Adv	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Gender												
Male	24	1.44	50	3	118	7.1	424	25.4	1053	63.1	1477	88.5
Female	15	1.59	21	2.22	64	6.8	215	22.7	631	66.7	846	89.4
Ethnicity												
Black, not Hispanic	12	2.03	25	4.23	41	6.9	146	24.7	367	62.1	513	86.8
White, not Hispanic	24	1.31	44	2.4	117	6.4	456	24.8	1196	65.1	1652	89.9
Not Reported: Native American or Alaska Native; Asian/Pacific Islander, and Hispanic groups*												
Primary Disability												
MR	20	1.58	28	2.21	84	6.6	327	25.8	806	63.7	1133	89.5
Not Reported: Specific LD, ED, Traumatic Brain Injury, Speech, Emotional, Hearing, Language, Visual, Orthopedic, Autism, Multiple Disabilities, and Other Health impairments*												
Student Status												
IEP	39	1.5	71	2.72	182	7	636	24.3	1680	64.4	2316	88.8
Not Reported: Gifted, H.S. Career Education, IAP, In district less than a year, In building less than a year, Migrant, Title 1,SES, and Voluntary Transfer Student designations*												
ELL Status												
Not Reported: Receiving ELL Services, ELL Monitoring, and Title III*												
Classroom Instruction												
From 21% to 60% of school day	14	1.83	16	2.1	37	4.9	174	22.8	522	68.4	696	91.2
More than 60% of school day	15	1.17	38	2.97	93	7.3	291	22.7	843	65.9	1134	88.5
Not Reported: Classroom Instruction Less than 21% of school day and Separate School *												

* In compliance with confidentiality requirements, data from these subgroups are not reported due to small sample size (*n* < 10 in any one cell).

Reliability and Validity

Validity refers to how well a test does the job it was employed to do. Reliability refers to the consistency of results from an assessment, or the extent to which an assessment provides the same results over repeated administrations and the extent to which various items within a test tend to provide the same results (AERA, 1999). The validity of any assessment is limited by its reliability. That is, if a test does not consistently yield the same results at each administration, it is probably not valid.

Reliability

Typically the reliability of assessments is determined by correlations among test-retest administrations, parallel forms, and items within the test (e.g., item discrimination, Cronbach's alpha). Neither parallel forms, test-retest reliability, nor consistency of an individual student's performance over time can be computed for the MAP-A as it is currently designed, administered, and scored. Recall that on each student's Entry/Data Summary Sheet there are six data points, three data points collected during each of two collection periods. These are averaged for a single entry score.

Internal consistency or homogeneity of the MAP-A can be computed as an estimate of reliability, with caution. Recall that two entries are completed for each of two strands within the mathematics or communication arts domains, and one entry is completed for each of four strands in science. Each entry assesses a single API. Thus, each student has four entry scores recorded for each of these domains. One measure of internal consistency, split-half reliability, is typically computed by dividing the test in half (e.g., odd vs. even items) and correlating scores on half the test items with scores on the other half. This approach could be used to estimate the reliability of the MAP-A in two ways:

1. Treat the two entries as two halves of a test and correlate the two scores. For mathematics and communication arts this would provide an estimate of internal reliability for each of the two strands.
2. Treat all four entries in mathematics, communication arts, or science as items of a test of the same domain and compute Cronbach's coefficient alpha.

Each API is supposed to represent the same strand, and each strand is supposed to represent the same domain. Thus, correlations between them provide an estimate of how generalizable each entry score is to the strand or to the larger domain. However, there are three concerns regarding the interpretation of these estimates:

1. This method depends upon variation among scores. The MAP-A has restricted variation. Teachers can select APIs and design assessment activities on which they are fairly certain the student will be successful. Thus, there is a negative skew on entry average scores, with roughly 51-66% of the scores at ceiling. The distribution of rubric scores is more restricted, with 72-90% scoring at ceiling and 6-12% scoring at floor, or "0."
2. This is a very short test. On the MAP-A, the split-half reliability would be based on only two or four items. The Spearman-Brown formula could be applied to estimate the reliability of the whole test if the test were twice as long (i.e., four or eight items), but even doubled it would be a short test. Reliability is a problem on a short test.
3. This method is best applied to similar items measuring a single concept. Ideally, the two halves of a test should have similar content and difficulty level. Items measuring each

behavior/skill should be on each half of the test. On the MAP-A, the halves are not likely to be equivalent because there is only one item on each half and because teachers are free to choose any two APIs from a field of dozens. For example, a 5th grader might be given the following two performance indicators: "*Recognize a small collection of 1 or 2 items"* (NO1.1a) and "*Develop fluency with basic number relationships of addition and subtraction for sums up to 10*" (NO9.4). Both of these APIs are designed to measure understanding of numbers and operations. However, they have different content and levels of difficulty.

Table 21 shows the domain of available APIs by content strand.

Table 21. 2015 Domain of Available and Assessed APIs

Grade	Strand	Total APIs Available	# of APIs Assessed
5	Scientific Inquiry **(IN)**	15	15
	Impact of Science, Technology and Human Activity **(ST)**	5	5
	Composition and Structure of the Universe and the Motion of the Objects within It **(UN)**	13	13
	Processes and Interactions of the Earth's Systems (Geosphere, Atmosphere, and Hydrosphere) **(ES)**	18	18
	Characteristics and Interactions of Living Organisms **(LO)**	19	17
	Changes in the Ecosystems and Interaction of Organisms with their Environments **(EC)**	14	12
8	Impact of Science, Technology and Human Activity **(ST)**	13	13
	Properties and Principles of Matter and Energy **(ME)**	38	31
	Processes and Interactions of the Earth's Systems (Geosphere, Atmosphere, and Hydrosphere) **(ES)**	38	33
	Scientific Inquiry **(IN)**	26	26
	Composition and Structure of the Universe and the Motion of the Objects within It **(UN)**	16	13
	Properties and Principles of Force and Motion **(FM)**	27	27
11	Scientific Inquiry **(IN)**	31	31
	Impact of Science, Technology and Human Activity **(ST)**	14	14
	Properties and Principles of Matter and Energy **(ME)**	54	35
	Properties and Principles of Force and Motion **(FM)**	37	28
	Characteristics and Interactions of Living Organisms **(LO)**	44	35
	Changes in Ecosystems and Interactions of Organisms with Their Environments **(EC)**	28	27

Tables 26 show the APIs that were assessed most often in content area.

Table 26. 2015 API Usage in Science

Grade	APIs Most Often Assessed	# of Times Assessed	% of Total Entries
Grade 5	ES7.1	308	8.58
	UN1.2	289	8.05
	IN5.1	194	5.41
	UN1.1	175	4.88
	LO1.3	145	4.04
	UN4.2	127	3.54
	EC1.5	121	3.37
	ES8.3	108	3.01
	ST1.3	96	2.68
	EC1.4	95	2.65
Grade 8	ME2.2	159	4.47
	FM1.2	152	4.27
	ME1.3	138	3.88
	ES7.1	134	3.76
	ME1.7	114	3.2
	ME1.1	94	2.62
	FM3.14	88	2.47
	FM3.1	76	2.13
	IN2.4	74	2.08
	ES7.2	80	2.02
Grade 11	EC2.3	184	5.56
	EC1.5	158	4.77
	EC1.4	127	3.38
	ME1.3	122	3.6
	LO1.3	106	3.2
	ME2.2	101	3.05
	LO1.4	83	2.51
	LO1.5	77	2.32
	ME1.7	74	2.23
	IN2.4	74	2.23

Noting these limitations to the interpretation of split-half reliability coefficients as applied to the MAP-A, Tables 27-30 report reliability estimates. Reliabilities for the rubric scores may be lower because the range is truncated.

Table 27. 2015 Reliability Estimates for the MAP-A, All Grades

		Science		
		Pair 1	Pair 2	Alpha
Entry Average	***n*=**	2066	2143	1792
Accuracy (0 – 100)		.80	.67	.85
Independence (0 – 100)		.81	.82	.90
Rubric Score	***n*=**	2615	2615	2615
Level of Accuracy (0 – 4)		.45	.49	.64
Level of Independence (0 – 4)		.50	.53	.67
Connections to Standards (0 – 3)		.47	.53	.67

Note. Numbers in the Strand 1 and Strand 2 columns present the Spearman-Brown split-half reliability coefficients for the two APIs within that strand. Alpha refers to Cronbach's alpha for the 4 API scores within each domain.

Table 28. 2015 Reliability Estimates for the MAP-A, by Grade Level

		Grade 5			Grade 8			Grade 11		
		Pair 1	Pair 2	Alpha	Pair 1	Pair 2	Alpha	Pair 1	Pair 2	Alpha
Entry Average	***n*=**	679	748	619	689	744	599	698	651	574
Accuracy (0 – 100)		.80	.66	.84	.77	.69	.84	.83	.67	.86
Independence (0 – 100)		.72	.80	.88	.81	.84	.88	.88	.81	.92
Rubric Score	***n*=**	897	897	897	890	890	890	828	828	828
Level of Accuracy (0 – 4)		.50	.58	.72	.37	.44	.55	.52	.45	.63
Level of Independence (0 – 4)		.51	.62	.73	.43	.49	.60	.58	.46	.67
Connections to Standards (0 – 3)		.48	.59	.70	.46	.49	.64	.50	.50	.66

Note. Numbers in the Strand 1 and Strand 2 columns present the Spearman-Brown split-half reliability coefficients for the two APIs within that strand. Alpha refers to Cronbach's alpha for the 4 API scores within each domain.

Three steps have been taken to increase the reliability of the MAP-A. First, three data points are collected at each of two collection periods for a total of six data points for each entry. The average for these six data points is taken as the student's score for that entry. Multiple data points result in a more stable score because the effects of "outlier" data points are minimized, and the average score is closer to what may be the student's "true" score. Increasing the number of data points should result in higher reliability.

Second, two standard forms, the "Entry/Data Summary Sheet" and the "Student Work Record," along with actual student work, if appropriate, are used to report data. Test administrators are carefully trained to provide data on these standardized forms. The degree of accuracy and of independence that is required to earn each point on the rating scales is clearly specified, and models are used in training. Data collection, documentation, and submission requirements are

prescribed in order to reduce the degree of variance in judgment that is somewhat inevitable in portfolio assessments. This standardized format contributes to reliability, although it has to be balanced with the need to design individualized assessments appropriate to each eligible student.

Third, scorers are carefully trained and monitored to assure inter-rater agreement. This is important because a test cannot have reliability that is higher than the reliability of the scoring. Inter-rater agreement is discussed in detail next.

Agreement Among Scorers

The extent to which two scorers assign the same score to an assessment when using the same rubric is referred to as inter-rater agreement. As part of ARC's quality control program for scoring MAP-A, inter-rater agreement reports are generated regularly. During scoring, 35% of submissions were given a blind second read. Thus, 915 of the 2014 MAP-A portfolios were checked for inter-rater agreement.

As a scorer completes a first read of a binder, his/her scores for each entry in the binder are entered into the MAP-A score database. As a scorer completes a second read of a binder, his/her scores for each entry in the binder are entered into the MAP-A score database and compared to the first set of scores. If there is a rubric score discrepancy on any of the entries within the portfolio, a facilitator then conducts a blind resolution read on the entry or entries in question. The facilitator's score then becomes the score of record.

Facilitators review discrepancy logs and agreement reports comparing inter-rater agreement percentages among scorers as well as agreement percentages with the facilitators' resolution reads. Early in the scoring season, agreement reports are reviewed several times a day with MAP-A program staff. As the season progresses and agreement rates stabilize, reports are reviewed by facilitators daily and with program staff several times a week.

Facilitators and program directors use inter-rater agreement and resolution reports to identify scorers in need of retraining and calibration and to identify any areas in which the entire scoring panel might have needed recalibration. With this information, retraining can be targeted and delivered quickly. Facilitators determine what retraining is necessary for scorers individually and as a group.

Table 33 summarize agreement reports for the MAP-A entries scored during the 2015 scoring season. Thirty-five percent of 10,460 science entries received second reads. Inter-rater agreement percentages for each subject may be found in the tables below. Level of accuracy and level of independence dimensions are scored using a four-point rubric. Connection to the standards is scored using a three-point rubric. The rubric for each scoring dimension calls for multiple decisions prior to assigning a rubric score. The maximum possible score per MAP-A entry is 11 points. The MAP-A scoring rules call for scorers to make decisions about whether an entry is scorable or unscorable. In cases of disagreement on such decisions, the resulting rubric scores differ by more than one point. This being the case, higher non-adjacent rates are expected in MAP-A scoring than in scoring using other holistic or analytic rubrics.

Table 33. 2015 Science Agreement Rates

	Perfect	Perfect Plus Adjacent	Non-adjacent
Level of Accuracy	97.62	99.59	0.41
Level of Independence	97.31	98.95	1.05
Connection to the Standards	93.89	95.91	4.09

Validity

Validity refers to the appropriateness, meaningfulness, and usefulness of inferences made from test scores. It is the extent to which an assessment measures what it is intended to measure for a particular purpose. The purposes of the MAP-A are to (1) document student learning according to state academic standards, and (2) inform instruction. Some of the evidence to support the validity of the MAP-A for these purposes has already been discussed in earlier sections of the report that address test administration, test scoring, and test reliability. Another important piece of evidence to support validity of the MAP-A for these purposes is test content, which is discussed next.

Test Content

Lissitz & Samuelsen (2007) argue that the test construction process is at the heart of validity. They state, "*content validity, or internal validity, should be acknowledged as the critical initial characteristic to consider when evaluating the quality of a test*" (p. 446). While there is controversy regarding whether test content is the most important aspect of validity (Embretson 2007), content validity is widely considered the minimal requirement for a valid test, but not a guarantee that a test is valid.

This aspect of validity refers to whether the content of the assessment corresponds with what content should be covered by the assessment, that is, whether test content is relevant and representative of the construct. It is based on judgment and is not quantifiable. We discuss three aspects of the MAP-A content that support its validity for the purposes discussed above:

1. The alignment of strands with standards;
2. The alignment of APIs with strands;
3. The range of content in portfolios.

First, during development of the MAP-A, a blueprint was used to outline the curriculum and standards for each subject and grade level. This process assured strong alignment of MAP-A strands with Missouri's Show-Me Standards, GLEs, and AGLEs. A summary of the assessment development process may be found in the Overview section of this report; refer to the *2006 MAP-A Technical Manual* for a detailed description of the mathematics and communication arts development process and to Appendix B for details regarding the science development process. The assessment blueprint may be found in the Operational Assessment Administration section.

Second, two steps have been taken to maximize alignment of APIs with strands. First, MAP-A administrators are carefully trained so that administration procedures are standardized. This process is described in the Operational Assessment Administration chapter. Second, each MAP-A portfolio is rated on its "connection to standards." This process is described in the Scoring and Reporting chapter. However, MAP-A administrators can choose what APIs to use to represent each strand with each student. Their choices influence the content validity of the MAP-A. In fact,

the validity of each student's portfolio is potentially unique, depending on the APIs selected by the administrator.

Third, effort has been made to broaden the range of content assessed by the MAP-A. Typically, tests merely sample a portion of the universe of items that could be used to assess a content domain. The larger the sample, the more valid the test. Because lengthy assessments are onerous, particularly for the MAP-A student, a balance must be achieved between the number of actual APIs selected and the universe of possible APIs. A 2006 study of communication arts and mathematics MAP-A submissions was conducted by Dr. Norman Webb, University of Wisconsin, at DESE's request, to address this issue.

Dr. Webb led an alignment study team using the Webb Alignment Tool (WAT), which has been used to analyze curriculum standards and assessments in over 16 states preparing to meet Title I compliance as required by the U.S. Department of Education. Overall, the findings from this study indicated need for improvement in the alignment between the collection of portfolios and the Missouri communication arts and mathematics alternate standards. Specifically, the MAP-A had limited range. Teachers were required to assess only two APIs for each of two strands in both communication arts and mathematics, yet there are a large number of APIs.

Although the state determined that the Webb model did not lend itself well to assessing the alignment of an alternate assessment of MAP-A's nature, DESE in 2008 took the following actions to improve alignment.

Teachers were provided with specific guidance in addition to the assessment blueprint, requiring them to select APIs not only from different strands, but also from different goals within the strands. To help teachers implement these new requirements, DESE provided additional training for teachers focusing on the following:

1. selection of APIs and design of activities at appropriate depth-of-knowledge levels, and
2. creation of assessment activities that closely tie to the content in the given APIs.

DESE provided for the development of additional sample entries and scoring information to be made available to teachers to assist them in their efforts to improve alignment.

Other states have used a variety of approaches to evaluating the alignment of alternate assessments, many based on modifications of the Webb model. DESE conducted a re-review of the mathematics and communication arts in conjunction with the NCLB-required alignment study of the science MAP-A, in 2009. This alignment study, conducted in collaboration with Human Resources Research Organization (HumRRO), used the Links for Academic Learning methodology, a significantly different approach designed specifically for alternate portfolio assessments. The technical reports for the alignment reviews of all MAP-A content areas can be found at http://dese.mo.gov/college-career-readiness/assessment/assessment-technical-support-materials.

Consequences of MAP-A Testing

The *intended* consequence of the MAP-A is to enhance education outcomes for children with disabilities. To this end reports are provided to parents, teachers, schools, districts, and DESE, as described in the Scoring and Reporting chapter. Achievement Level Descriptors (ALDs) provide users with clear reference points for mastery at each grade level, so that scores can be readily interpreted and used to inform curriculum and IEP development. However, different APIs are used from year to year, so annual growth for individual children for specific APIs cannot be tracked.

Assessments can also have both positive and negative *unintended* consequences. Researchers disagree about whether assessment of consequences is an aspect of validity of a test or not, but there is widespread agreement that test designers and users should explore and fully disclose identified consequences of a test's use, including negative consequences, whenever possible (Linn 1997; Popham 1997; Shepard 1997).

Therefore, DESE commissioned a study to evaluate the consequences of its state assessment program. Part of that study addressed the consequences of MAP-A. Focus group discussions and surveys were used to collect information from several stakeholder groups, among them teachers, parents, students, school board members, superintendents, principals, and personnel from DESE, and its Regional Professional Development Centers. Through this study and other contact with MAP-A stakeholders, a number of findings have emerged, both positive and negative.

1. MAP-A design lends itself to incorporation into IEP goals.
2. Requirements to administer the assessments led to better interventions for some MAP-A students.
3. MAP-A documentation and time requirements are onerous.
4. It is difficult to select appropriate APIs for the most severely disabled students.
5. Teachers' knowledge or lack of knowledge about how to administer the assessment and about the content standards affects student scores.

These findings suggest that stakeholders perceive the MAP-A as valid for the purpose of informing instruction. The findings also suggest that the assessment is challenging for teachers. Findings from multiple perspectives were presented in a symposium at the American Educational Research Association's annual meeting in April 2009.

Teachers' Role

Teachers have a significant role in administering, reporting, and using the information provided by the MAP-A. Thus, teachers influence the validity of the test. DESE provides training and on-going guidance to help teachers administer and report the assessment validly. Nevertheless, teachers introduce construct-irrelevant variance that may compromise the validity of the MAP-A. There are three ways that administration error can reduce a student's score.

1. If a teacher fails to provide evidence of evaluation on a student work record, the student would get a "0" on the accuracy and independence scores for that data point. This "0" would be averaged with the other two data points for that collection period. (If the teacher miscalculates, the entry is simply re-calculated, which could lead to a lower or higher score.) Thus, a student who may be fully capable of an API, but whose teacher fails to adequately document this on the student work record, would get a score of "67" [(100 +

100 + 0)/3] instead of a score of "100." This would result in a lower rubric score, and may or may not result in a lower overall achievement level.

2. If a teacher gives the student an *acquisition* rather than *application* task, the student would get a lower "connections to standards" score, which would reduce the rubric score to 9-10 instead of 11. This may or may not result in a lower overall achievement level.
3. If a teacher (a) chooses an API not in the grade span, (b) describes an activity that doesn't connect with the API, or (c) assesses the student outside the specified time period, the student would receive a "no score" for that API, which becomes a "0" for the rubric score. For example, the API that "Cody" was assessed on was "*Write simple directions for doing something, considering a given audience*" (WP5.4). Cody wrote a grocery list for a recipe to be prepared by his life skills class. Cody showed accuracy and independence, but received a rubric score of "0" because his teacher simply reported that Cody found the ingredients, but did not discuss his writing, nor what kind of prompt was needed. Cody's score of "0" suggests inability to complete this API, when in fact he could write a shopping list. A rubric score of "0" would reduce his overall score by 11 points, out of a possible 44. This is likely to place him in a lower overall achievement level.

Teacher error in administration of the MAP-A could result in artificially low scores for students, whereas a correct administration could have permitted the students to display their competence. Thus, the meaning of a particular student's rubric score is not entirely clear, and may or may not be valid for determining the student's overall achievement level.

In summary, we cannot know all aspects of validity and reliability of the MAP-A because of the nature of this assessment. We cannot compare scores from one student to another. We cannot know how their performance pertains to same-age peers who are completing standardized assessments. However, strong efforts have been made to ensure that the assessment is as valid and reliable as possible for an individualized performance assessment. The evidence described above suggests that the MAP-A's psychometric properties contribute to its intended consequence, that is, to make inferences about student achievement on the Show-Me Standards for communication arts, mathematics, and science and to improve instructional programs.

MAP-A Information Security

Although the MAP-A submissions do not contain secure test items, they do contain confidential student information. The security of this information is maintained throughout the MAP-A cycle, from enrollment to receipt and check-in of submissions and through scoring, reporting, and archiving.

Enrollment

Electronic enrollment is handled by an ASP.NET website with a back-end Oracle database located behind a firewall. The website is protected by 128-bit SSL encryption, and the webserver is protected with IP filters for minimal exposure. The website requires users to login with a username and password assigned by ARC. District test coordinators can elect to create accounts within the system that can be used by their designees to enroll students. Enrollment is limited to students within a district and edit/delete can only be done by the district test coordinator.

Scoring

MAP-A binders returned to ARC for scoring are shipped to and stored in a secure warehouse adjacent to the rooms where scoring takes place. Access to the warehouse is limited to employees of ARC. Binders are staged for scoring in a secure manner. All ARC staff, including scoring personnel, sign a confidentiality agreement that is legally binding in which they agree not to discuss any aspect of the scoring process or confidential student information. The scoring process and confidential student information are defined to include, but not be limited to, any aspect of scoring, student responses, districts or teachers administering the MAP-A outside the scoring room. In addition, all ARC staff wear security identification name badges at all times during the workday. No cell phones, cameras, or other recording devices are allowed in scoring areas. All materials necessary for scoring, including training materials, rubrics, and MAP-A binders, remain in designated scoring areas. When scoring is concluded, discarded paper and scoring materials are securely shredded.

Data Storage

The enrollment data and score data are stored on University of Missouri servers which are behind firewalls. Additional network-level protection is provided by IP filters that block access to unauthorized subnets and protocols, regardless of their presence inside the intranet. Data are stored in a combination of Oracle database and flat text file formats. File-level access control lists prevent unauthorized staff from accessing MAP-A data on the network.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing.* Washington DC: American Educational Research Association.

Bergin, D. A., Bryant, R. A., McFarling, P. L., Murphy, B. E., Parshall, T., Sireno, L., Su, I. (2009). Motivational Aspects of NCLB-Mandated Testing. Presentation at the American Educational Research Association, April 2009, San Diego, Symposium: Intended and Unintended Consequences of NCLB-Mandated Testing.

Bryant, R. A., Murphy, B. E., Bergin, D. A., McFarling, P. L., Parshall, T., Sireno, L., Wang, Z. (2009). Perceptions of Responsibility and Accountability for Student Learning in the Context of NCLB-Mandated Testing. Presentation at the American Educational Research Association, April 2009, San Diego, Symposium: Intended and Unintended Consequences of NCLB-Mandated Testing.

Embretson, S. E. (2007). Construct validity: A universal validity system or just another test evaluation procedure? *Educational Researcher* 36(8), 449–455.

Linn, R. L. (1997). Evaluating the validity of assessments: The consequences of use. *Educational Measurement: Issues and Practice* 16(2), 14–16.

Lissitz, R. W., & Samuelsen, K. (2007). A suggested change in terminology and emphasis regarding validity and education. *Educational Researcher* 36(8), 437–448.

McFarling, P. L., Bryant, R. A., Parshall, T., Sireno, L. (2009). Overview of the Missouri Assessment Program and Missouri Schools Study. Presentation at the American Educational Research Association, April 2009, San Diego, Symposium: Intended and Unintended Consequences of NCLB-Mandated Testing.

Popham, W. J. (1997). Consequential validity: Right concern - wrong concept. *Educational Measurement: Issues and Practice* 16(2), 9–13.

Shepard, L. A. (1997). The centrality of test use and consequences for test validity. *Educational Measurement: Issues and Practice* 16(2), 5–24.